

What is claimed is:

1. A method comprising:

detecting a phone in an off-hook state;

5 retrieving with a telephony server information associated with a user assigned to the phone;

generating a custom input grammar with the telephony server using the information;

generating a dial-tone with the telephony server;

receiving with the telephony server a command spoken into the phone;

10 processing the spoken command with the telephony server to locate a corresponding entry in the custom input grammar; and

executing a command operation associated with the corresponding entry.

2. The method of claim 1, wherein the custom input grammar is not generated until an

15 identification of a person who spoke the command is performed, and wherein the custom input grammar is then generated based on the particular profile of the person.

3. The method of claim 1, wherein said processing comprises:

sending the spoken command to a speech recognition server, said speech recognition

20 server processing the spoken command, locating the corresponding entry in the custom input grammar, and returning the corresponding entry to the telephony server.

4. The method of claim 3, wherein the speech recognition server verifies the identity of a person that spoke the command to ensure the person is authorized to access the custom input grammar before locating the corresponding entry in the custom input grammar.

5 5. The method of claim 1, wherein the custom input grammar is generated from a text-based contacts database associated with the user assigned to the phone.

6. The method of claim 5, wherein the spoken command is a name of a person in the text-based contacts database associated with the user assigned to the phone.

10

7. The method of claim 1, wherein the command is spoken into the phone by a person other than the user assigned to the phone.

8. The method of claim 1, wherein said generating the custom input grammar is only
15 performed if the custom input grammar does not already exist for the user associated with the phone or if the custom grammar exists but needs updated due to modifications in an underlying data source.

9. The method of claim 1, wherein the dial-tone is cancelled when the telephony
20 application processor begins receiving the command spoken into the phone.

10. The method of claim 1, wherein said processing the spoken command with the telephony application processor comprises:

sending a recognition request to a speech recognition server;

receiving a probing request from the speech recognition server;

5 sending a UDP probe response message to a probing port number of the speech recognition server;

sending the spoken command to the speech recognition server, said speech recognition server determining a translated result based on the custom input grammar; and

receiving the translated result from the speech recognition server.

10 .

11. A method comprising:

providing a probing endpoint for a first server;

receiving at a second server a port number of the probing endpoint of the first server;

15 receiving at the second server a delivery request for which probing is requested from the first server; and

sending a UDP probe response message to the port number of the first server.

12. A method comprising:

20 providing a probing endpoint for a speech recognition server;

receiving at a telephony server a port number of the probing endpoint of the speech recognition server;

receiving at the telephony server an audio delivery request for which probing is requested from the speech recognition server; and

sending a UDP probe response message to the port number of the speech recognition server.

5

13. The method of claim 12, further comprising:

receiving the probe response message at the speech recognition server.

14. The method of claim 13, further comprising:

10 reviewing an identifier contained in the probe response message to confirm the probe response message was received from the requested telephony server.

15. The method of claim 12, further comprising:

sending at least one audio packet to the speech recognition server.

15

16. The method of claim 15, wherein the at least one audio packet sent to the speech recognition server was a command spoken into a phone and received by the telephony server.

17. The method of claim 12, wherein said receiving at the telephony server the port

20 number of the probing endpoint occurs when the telephony server first requests a recognition operation from the speech recognition server.

18. The method of claim 12, wherein the probing endpoint has a same IP address as used for delivering audio to the speech recognition server.

19. A method comprising:

5 providing a probing endpoint for a speech recognition server; and
sending from the speech recognition server a plurality of probing requests to a telephony server until the telephony server sends a UDP probe response message or until a predetermined quantity of missed probes has been exceeded.

10 20. A method comprising:

providing an audio streaming packet;

receiving an RTP physical sequence number associated with the streaming audio packet;

receiving a last logical sequence number that was most recently generated; and

15 generating a new logical sequence number by a process comprising:

adding a fixed-size kilobyte amount to the RTP physical sequence number;

generating a scale factor by subtracting the fixed-size kilobyte amount from the last logical sequence number and masking off from the result a plurality of bits from a lowest bit range; and

20 adding the scale factor to the RTP physical sequence number.

21. The method of claim 20, further comprising:

adjusting the new logical sequence number using a revised scale factor if the new logical sequence number is not within a certain predetermined range of the last logical sequence number.

- 5 22. The method of claim 20, wherein the process repeats for each of a plurality of audio streaming packets received.

23. The method of claim 22, wherein the new logical sequence number of each of the plurality of audio streaming packets is used to reassemble the plurality of packets into a
10 logical order.

24. A method comprising:

allocating an internal buffer list with a plurality of fixed size buffers totaling a maximum receive packet size;

- 15 passing the internal buffer list to an operating system as a scatter/gather array;
filling at least a portion of the plurality of fixed size buffers in order when a packet is received; and

freeing the unused fixed size buffers back to the internal buffer list.

- 20 25. The method of claim 24, wherein the packet is a streaming audio packet.

26. The method of claim 25, wherein the streaming audio packet contains information to be processed by a speech recognition server.

27. The method of claim 26, wherein the information to be processed by the speech
5 recognition server includes a command that was spoken into a phone.

28. A system comprising:
a speech recognition server; and
a telephony application server coupled to the speech recognition server over a
10 network, the telephony application server being operative to detect a phone in an off-hook
state, retrieve information associated with a user assigned to the phone, generate a custom
input grammar using the information, generate a dial-tone, receive a command spoken into the
phone, send the spoken command to the speech recognition server, receive a corresponding
entry based on the custom input grammar from the speech recognition server and execute a
15 command operation associated with the corresponding entry.

29. The system of claim 28, wherein the speech recognition server is operative to support
a plurality of speech recognition engines.

20 30. The system of claim 28, wherein the speech recognition server is operative to send a
port number of a probing endpoint to the telephony application server, send a probing request

to the telephony application server, and receive from the telephony application server a UDP probe response message at the port number.

31. A method comprising:

- 5 installing a particular speech recognition engine ;
establishing grammar for the particular speech recognition engine after said installing;
installing a speech recognition subsystem on a telephony application server after said
establishing, the speech recognition subsystem including an application interface operable
with multiple speech recognition engines, two or more of the multiple speech recognition
10 engines being incompatible with one another and the multiple speech recognition engines
including the particular speech recognition engine previously installed; and
operating the telephony application server with the grammar from said establishing.

32. The method of claim 31, wherein the speech recognition server and the application

15 server are the same server.

33. The method of claim 31, wherein a different speech recognition engine is used by
modifying an identifier that specifies which grammar syntax is to be used by the different
engine.

20 34. A system comprising:

multiple speech recognition engines residing on one or more speech recognition
servers; and

a telephony server having a telephony application processor operable to translate vendor-neutral interfaces to and from a specific syntax required by each of the multiple recognition engines.

5 35. The system of claim 34, wherein the telephony application processor is operable to perform speaker identification and verification as part of a recognition operation.

36. The system of claim 34, wherein the telephony application processor is operable to send recognition requests to at least two of the multiple speech recognition engines at the
10 same time.

37. A method, comprising:

 offering a telephony application interface routine including a voice recognition interface operable with multiple speech recognition engines;

15 providing the telephony application interface to a first customer having a pre-established grammar for a first one of the speech recognition engines;

 the first customer operating the telephony application interface with the pre-established grammar of the first one of the speech recognition engines;

 providing the telephony application interface to a second customer having a second
20 one of the speech recognition engines; and

 the second customer operating the telephony application interface with the second one of the speech recognition engines.

38. A method, comprising:

operating a telephony application interface routine including a voice recognition interface operable with multiple speech recognition engines, said operating including

5 interfacing with a first one of the speech recognition engines;

obtaining a second one of the speech recognition engines; and

interfacing the telephony application interface routine with the second one of the speech recognition engines.

10 39. A method comprising:

detecting a user being connected to a telephony server;

identifying the user;

retrieving information associated with the user;

generating a custom input grammar using the information;

15 receiving with the telephony server a command spoken by the user;

processing the spoken command to locate a corresponding entry in the custom input grammar; and

executing a command operation associated with the corresponding entry.

20